

## ¿HAY ALGUIEN AHÍ? SOBRE LOS FILÓSOFOS Y LAS MÁQUINAS

(Nueva Revista, 4, V-1990, pp 41-45)

Hay ciertas intrigantes cuestiones que acompañan al hombre desde tanto tiempo que suelen llamarse eternas. Lo que muchas de ellas tienen de desconcertante es el contraste entre la facilidad con que pueden ser formuladas -y entendidas incluso, y a veces sobre todo, por un niño- y la enorme dificultad para contestarlas de un modo mínimamente coherente. Un ejemplo especialmente obvio de ese tipo de preguntas es el siguiente: "¿Qué significa pensar?"

Dicha pregunta no sólo es un caso particular del tipo de inquietudes que asaltan en momentos de turbación y perplejidad al ciudadano medio (y que se supone ocupan el tiempo de los filósofos) sino que tiene la peculiaridad de que se suscita con cierta facilidad cuando tratamos de estar al día en los progresos de la Inteligencia artificial y materias conexas. En realidad, semejante cuestión es tanto una pregunta con pasado como una interrogación que inevitablemente nos planteará el futuro.

Un reciente libro de Roger Penrose -un importante físico y matemático inglés, coautor con el famosísimo Hawking de algunas notables aportaciones a la Cosmología actual- tiene la extraordinaria virtud de examinar esta cuestión con el debido espesor y con la mayor claridad que un empeño tan arduo permite.

La exposición de Penrose tiene, junto a la virtud de su originalidad, el atractivo adicional de cierta heterodoxia filosófica; su libro podría haber sido escrito por un clásico: la ortodoxia del momento es muy circunspecta respecto a lo que se puede decir y a lo que debemos callar. Para valorar la aportación del autor, parece conveniente situar el problema en sus coordenadas fundamentales.

### 1. Historia de una ilusión

Desde los tiempos de Leibniz -y desde los aún más lejanos del mallorquín Raimundo Lulio- los pensadores de orientación matemática han acariciado la idea de encontrar un "ars combinatoria" que pudiera resolver sus dudas y cavilaciones, un sistema que permitiera reducir la ignorancia mediante el cálculo.

"Calculemos", pensaba Leibniz que se podría decir una vez construida esa maravilla. Toda verdad podría ser deducida y demostrada sin posibilidad de error ni de duda: la mente del hombre se vería así liberada de la servidumbre del error, de su unión a un cuerpo que, como ya viera Descartes, era la auténtica causa de sus desvaríos.

Una versión de esta clase de sueño es la convicción contemporánea de que se pueden construir máquinas que piensen, artefactos que pueden ser vistos como "mentes" al menos con tanta propiedad como las personas pueden serlo.

La pregunta por el pensar se refiere ahora a la máquina. A.M. Turing, uno de los fundadores de la inteligencia artificial, analizó en un célebre artículo de 1950 cuales habrían de ser los criterios que llevaran a decidir la respuesta válida a la pregunta de si una máquina puede pensar. El

llamado Test de Turing es la concreción práctica de dichos criterios: podremos decir que una máquina piensa cuando sea capaz de mantener una conversación con una persona, sin que ésta pueda descubrir que está hablando con una máquina.

El desarrollo de la inteligencia artificial no se ha orientado a fabricar sistemas capaces de superar el Test de Turing entre otras razones porque, como en toda tecnología, Los diseños de la investigación han sido más prácticos que filosóficos. Por lo demás, el test podría ser superado, en ciertas condiciones, por programas relativamente simples -por ejemplo, por aquellos que imitan la conversación de algunos psicoterapeutas- y no podría ser ni siquiera planteado si exigiéramos a la máquina algo más sustancial que un diálogo de compromiso.

La bibliografía sobre si las máquinas piensan es ingente y continua creciendo. Tan sólo hace unos meses, "Scientific American" (y su edición española "Investigación y Ciencia") recogía una polémica sobre el particular: el filósofo J. R. Searle argumenta en contra de la identificación entre mentes y programas (amparándose en la distinción entre sintaxis y semántica) mientras que Paul B. Churchland y Patricia Smith Churchland sostienen lo contrario (las máquinas pueden pensar, y lo acabarán haciendo, especialmente aquellas que imiten el modo de funcionamiento - en paralelo y no en serie- del cerebro humano).

Sobre esta cuestión de la mente gravita buena parte de la historia del pensamiento desde Parménides y Platón: no es extraño que muchos autores hayan sucumbido a las dificultades que plantea y la den por imposible. Tampoco debiera extrañarse nadie de que abunden las rectificaciones y los cambios de opinión. El filósofo Hilary Putnam, que ha dedicado al tema varias y valiosas contribuciones, advierte en el prólogo de un libro de 1988 que va a criticar algunas opiniones corrientemente sostenidas, entre otros, por él mismo.

¿Puede pensar una máquina? La respuesta depende, entre otras cosas de qué entendamos por "pensar" y por "máquina". Si partimos de un concepto de máquina tan amplio, que el hombre pueda ser considerado como una "maquina", entonces, evidentemente, hay máquinas que piensan. Si, por otro lado, llamamos "pensar", simplemente, a manipular símbolo, entonces, hasta una calculadora de bolsillo es un notable caso de pensador.

Las dificultades se presentan solo cuando las definiciones de "pensamiento" y "maquina" son más estrictas, cuando se acercan más a las intuiciones comunes que sobre ellas tenemos.

## **2. Los argumentos**

Los hombres, sin duda, piensan. Pero ¿piensan con sus mentes o con sus cerebros? ¿pensar es una actividad que requiera instrumentos? ¿se "piensa-con" o simplemente "se piensa"? Podríamos multiplicar los interrogantes sin ninguna dificultad pero, también, sin mucha ganancia, puesto que lo que se discute es cuales son las preguntas pertinentes. El primer problema que se ha de afrontar, y que muchas veces se olvida, es, por supuesto, el que plantea el pensamiento humano. Las doctrinas al respecto son el dualismo, que no está de moda (precisamente porque fue "doctrina oficial" en un pasado relativamente reciente), y el materialismo. Los dualistas mantienen que la mente humana es algo distinto del cerebro y que interactúa con él (en formas y modos difíciles, cuando menos, de precisar), y los materialistas opinan que mente y cerebro son una y la misma cosa (con distintas apariencias debidas a nuestro modo de conocer e, incluso, a nuestro modo de hablar).

Los dualistas que no creen que el cerebro piense se sienten muy poco inclinados a conceder que las máquinas lo hagan. Los materialistas están, al respecto, más divididos: los que creen que lo que hace el cerebro pensar es su carácter biológico dudan de que una "máquina" no fisiológica pueda llegar a nada semejante. Los que creen que "pensar" es algo que los cerebros hacen, pero algo que también podrían hacer otras entidades, creen que las máquinas llegarán a pensar cuando sean suficientemente complejas.

El segundo problema, que también se olvida en ocasiones, es, en cierto modo, más peliagudo aún que el primero: ¿cómo sé yo que alguien piensa? ¿cómo sé que hace o siente o "ve" aquello que yo hago, siento y "veo" cuando digo que pienso? Yo supongo que mis semejantes piensan porque hacen cosas análogas a las que yo hago cuando pienso; puesto que no tengo experiencia inmediata del pensamiento ajeno, o bien hago caso del "Test de Turing" e identifico pensamiento con "conducta inteligente", o bien separo una cosa de otra y entonces puedo llegar a suponer que tal vez hasta las piedras piensen. ¿Qué clase de conciencia tienen los animales? ¿Cómo es ser un murciélago?, son algunas de las preguntas que la dificultad que puede llevar al Solipsismo obliga a plantear. Así pues, careciendo de criterios para decidir, en base a la experiencia, si algún sujeto piensa o no ¿qué decir acerca de si piensan o no las máquinas?

Un tercer problema, siempre presente, es el de la relación entre "conciencia" y "pensamiento" o, dicho de otro modo, el de la relación entre la conciencia y el argumento de un pensamiento en tanto puede ser objetivado (escrito, comunicado, demostrado por una "maquina" etc.). Una última dificultad, es la planteada por la relación entre el ser consciente y su constitución propia en cuanto parte de un mundo o en cuanto presente, como tal, ante otros seres conscientes (que observan no la inaccesible conciencia, sino la inmediata corporalidad).

Las distintas doctrinas sobre el pensamiento abordan los problemas anteriores (y una gran variedad de cuestiones derivadas) con diversas estrategias. A riesgo de simplificar (lo que es siempre excesivo) se puede decir que tales estrategias dependen, en último término, de la preferencia por alguno de los siguientes argumentos, de su jerarquía y de la interpretación que de ellos se haga:

- a. La conciencia es una propiedad natural que la evolución ha ido perfeccionando en la escala animal asociándola a la complejidad y sofisticación del cerebro.  
La conciencia debe ser distinguida de toda pretensión de formalizar cualquier esquema válido de conocimiento.
- b. La conciencia en tanto posee propiedades intencionales (esto es, se refiere a objetos que parecen independientes de ella) es irreductible al tipo de propiedades comunes a los objetos físicos en cualquier forma que éstos se le presenten.
- c. La conciencia es libre respecto a las leyes que puede descubrir en el mundo de los objetos que conoce.
- d. La conciencia es el resultado de la interacción entre ciertas formas de realidad cuyo estudio corresponde por un lado a la Física y la Fisiología (relación mundo-cerebro) y por otro a la Lógica y la Psicología (relación entre subsistemas cerebrales y núcleos de conceptos y perceptos).

Aceptar o rechazar estos argumentos, en las distintas formas en que puede hacerse, da lugar a todo el surtido de teorías de la mente con alguna vigencia y determina la respuesta a la cuestión sobre pensamiento y máquinas.

### 3. La posición de Penrose

Nuestro autor aborda la cuestión por sí solo y desde fuera de la controversia filosófica tal como ahora está planteada. Además, Roger Penrose es un físico y no precisamente un físico muy convencional: está acostumbrado a vérselas con cuestiones harto difíciles en Cosmología y en Mecánica cuántica y a sugerir hipótesis tan originales y atrevidas como su teoría del "twistor" o los "teoremas de la singularidad".

Su obra es, en cierto modo, un homenaje a las preguntas profundas de los niños que aún no han sucumbido a la ruda disciplina de los "pseudoproblemas". El prólogo y el epílogo insisten en ello: sin duda hay cuestiones profundas que pueden hacer reír a los que se tienen por sabios. Su libro es, también, una reacción frente a las suposiciones de algunos de los "profetas" más atrevidos de la inteligencia artificial.

El texto está lleno de agudas observaciones (por ejemplo: tal vez los filósofos que sólo usan palabras en sus discursos tienden a exagerar el papel del lenguaje en el pensamiento) y no puede ser, ni de lejos, resumido en unas pocas palabras (es obvio que, a veces, *sí* se puede resumir un libro de 500 páginas).

Penrose comienza por analizar el problema en los claros términos escogidos por Turing; a partir de ahí, va tirando de un hilo que atraviesa por la mayor parte de las más abruptas regiones del conocimiento humano: qué es y qué no es un algoritmo, qué papel representa en el conocimiento matemático, cuales son los fundamentos de éste; cuál es la naturaleza de la realidad física; como es la neurofisiología del cerebro, qué tiene que ver con los modelos de ordenador; Penrose está convencido de que nuestro conocimiento del cerebro y de cómo funciona la conciencia depende en muy buena medida de esas y otras cuestiones de la Física, la Psicología, y la Filosofía.

Así, repasa los fundamentos de la Mecánica cuántica, su estructura y sus enigmas, las dimensiones principales de la Física clásica y relativista, de la Termodinámica y, por extraño que resulte, las andanzas del "Big bang" y los "agujeros negros".

Nada sería tan fácil como perderse en una excursión tan complicada; pero Penrose tiene muy claro lo que busca y en ningún momento abandona el hilo de Ariadna. El lector puede seguirle y, si en algún caso no lo hace, está autorizado (por el autor) a pegar un pequeño salto hasta encontrar de nuevo la senda.

¿Cuáles son las aportaciones de Penrose? Hay que insistir en la necesidad de seguir el libro para valorar la novedad y la coherencia de sus planteamientos; sin embargo hay unas cuantas nociones que Penrose introduce y que son susceptibles de análisis aislado; veámoslas:

En primer lugar, Penrose mantiene que el Test de Turing proporciona tan sólo una evidencia indirecta de lo que es la inteligencia aunque rechace la idea de que, por el solo hecho de que el computador esté construido con componentes inorgánicos, le sea imposible ser consciente; pero el Test de Turing no es un detector suficiente de la conciencia, según Penrose.

Tras examinar el argumento anti-Turing de Searle (un experimento mental conocido como "la habitación china") avanza una nueva opinión: ningún algoritmo, por complicado que sea, incluye por sí mismo genuino entendimiento; ni los algoritmos tienen mucho que ver con el

conocimiento matemático, ni está claro que lo que los cerebros hacen sea ejecutar algoritmos. Penrose combate la presunción de que "todo es un computador digital" que se asienta en una deficiente apreciación de las propiedades que necesariamente ha de tener el "hardware" y en la supuesta equivalencia funcional de todas las máquinas de este tipo.

Penrose entiende, como también el último Putnam, que el Teorema de Gödel (que acabó con la pretensión de formalizar y mecanizar toda la matemática) evidencia que la inteligencia humana trasciende toda fórmula concebible. Para Penrose, la conciencia tiene un ingrediente no algorítmico que se hace patente en la distinción entre intuición y cálculo de la Matemática. Penrose, que se confiesa geómetra, se refugia en la "objetividad platónica" para sugerir lo que es propiamente la inteligencia matemática.

Llegados a este punto se puede situar la principal novedad del análisis de nuestro autor: el ingrediente no algorítmico de la conciencia humana se debe a que el cerebro "utiliza" la física a un nivel cuántico y no, como suele suponerse, en un nivel meramente macrofísico y determinista. Esto hace que el cerebro no sea en realidad un ordenador sino más bien un ordenador que cambia de modo continuo su propia constitución física y su propia red de comunicaciones. Así, la mente consciente no puede ser un simple (por complicado que sea) programa aunque se sirva de muchos de ellos.

La sugerencia de Penrose no es nueva: lo nuevo es el modo de hacerla; muchos de sus argumentos son especulativos aunque otros son absolutamente incontestables. Así, entre demostraciones y especulaciones, se desarrolla la jornada penrosiana.

El libro consigue, en muy amplia medida, lo que motivó su escritura: mostrar que se necesita una teoría de la conciencia, y que la que por tal tenemos, es notoriamente insuficiente.

#### **4. A modo de conclusión**

Buena parte de los filósofos profesionales practica hoy una pudibunda ascética que les impide afrontar las viejas cuestiones en todo su selvático esplendor. Parece que el miedo a equivocarse inhibe la audacia intelectual hasta la más tautológica esterilidad: uno de los más temibles síndromes de esta curiosa dolencia es el del "rechazo del dualismo".

Penrose no ha debido oír hablar de esos peligros, y si ha oído hablar de ellos no se los cree. El caso es que tanto su visión de lo que es la conciencia -algo científicamente describable, que no es puramente pasivo, sino que activamente "entiende"-, como su idea de lo que principalmente hace -formación de juicios- como, por último, sus hipótesis acerca de como se relaciona todo ello con nuestro conocimiento de la naturaleza física, habrían hecho sonreír complacientemente al viejo Descartes, cuyo oficio principal comparte Penrose.

¿Cabe esperar que todo ello sirva para reavivar el debate sobre estas cuestiones y para bajar los humos a algunos materialistas un poco apresurados? Contestar afirmativamente, supondría dominar las "variables ocultas" de la dinámica de la cultura asunto casi tan difícil de suyo como la propia, Mecánica cuántica o la mismísima teoría de la mente. "Habent sua fata libelli" gustaba de repetir Ortega, aunque no se debería seguir hablando de estas cosas sin leer a Penrose.